Leveraging Attention Mechanisms for Interpretable Human Embryo Image Segmentation

Wided Souid Miled[1,2] and Nozha Chakroun[3]
[1]*LIMTIC Laboratory, Higher Institute of Computer Science, University of Tunis El-Manar, Ariana, Tunisia*
[2]*National Institute of Applied Science and Technology, University of Carthage, Centre Urbain Nord, Tunisia*
[3]*University of Medicine of Tunis, Lab. of Reproductive Biology and Cytogenetic, Aziza Othmana Hospital, Tunisia*

Keywords: In Vitro Fertilization, Embryo Selection, Blastocyst Segmentation, Attention Mechanisms, Deep Learning, Explainability.

Abstract: In-vitro Fertilization (IVF) is a widely used assisted reproductive technology where embryos are cultured under controlled laboratory conditions. The selection of a high-quality blastocyst, typically reached five days after fertilization, is crucial to the success of the IVF procedure. Therefore, evaluating embryo quality at this stage is essential to optimize IVF outcomes. Advances in neural network architectures, particularly Convolutional Neural Networks (CNNs), have enhanced decision-making in IVF. However, ensuring both accuracy and interpretability in these models remains a challenge. This paper focuses on improving human blastocyst segmentation by combining channel attention mechanisms with a ResNet50 model within an encoder-decoder architecture. The method accurately identifies key blastocyst components such as inner cell mass (ICM), trophectoderm (TE), and zona pellucida (ZP). Our approach was validated on a publicly available human embryo dataset, achieving Intersection over Union (IoU) scores of 83.09% for ICM, 86.87% for ZP, and 81.1% for TE, outperforming current state-of-the-art methods. These results demonstrate the potential of deep learning to improve both accuracy and interpretability in embryo quality assessment.

1 INTRODUCTION

In vitro fertilization (IVF) is one of the most widely used and effective forms of assisted reproductive technology (ART) that helps couples facing fertility challenges conceive a child. Since 1978, more than 9 million babies have been born through IVF, and approximately 6% of couples experiencing infertility turn to this procedure (Kuhnt and Passet-Wittig, 2022). The IVF process involves several critical stages. First, mature eggs are retrieved from the ovaries and manually combined with sperm in a controlled environment for fertilization. The resulting fertilized egg, now called embryo, undergoes a series of developmental phases. Initially, the male and female pronuclei appear and then disappear, followed by the cleavage stage, where the single cell divides into multiple cells. Four days after fertilization, the embryo compacts, reaching the morula stage, and by the fifth day, the embryo develops into a blastocyst.

One of the most critical steps in this complex process is embryo selection, which aims to identify the healthiest embryo with the highest likelihood of resulting in the birth of a healthy baby. Embryo quality is considered a key predictor of success in IVF cycles. Numerous studies have demonstrated a strong correlation between embryo morphology, implantation rates, and clinical pregnancy outcomes (Shulman et al., 1993; Dennis et al., 2006). Consequently, selecting a high-quality embryo significantly increases the potential for a successful pregnancy. A noteworthy advancement in the field of IVF is the introduction of time-lapse imaging incubators (TLI). This innovative technology has transformed the embryo selection process by providing a dynamic, real-time view of embryonic development. TLI systems capture images of each embryo at regular intervals and compile them into a time-lapse video, offering dynamic insight into embryonic development in vitro without disturbing the stable culture conditions (Kovacs, 2016; Goodman et al., 2016). Using these time-lapse videos, embryologists grade blastocysts and identify the embryos with the greatest pregnancy potential for transfer to the woman's uterus.

Miled, W. S. and Chakroun, N.
Leveraging Attention Mechanisms for Interpretable Human Embryo Image Segmentation.
DOI: 10.5220/0013383200003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 2*, pages 872-879
ISBN: 978-989-758-737-5; ISSN: 2184-433X

At the blastocyst stage, the embryo consists of two main inner regions, the inner cell mass (ICM), and the trophectoderm epithelium (TE). These regions are surrounded by an outer membrane, the zona pellucida (ZP). Both ICM and TE are considered key morphological parameters for assessing embryo viability. Therefore, evaluating the quality of these regions is essential to determine embryonic potential. The most commonly employed grading system for blastocyst morphology is that of Gardner *et al.* (Gardner and Schoolcraft, 1999). According to this system, three crucial parameters are used to predict successful pregnancy outcomes: the degree of blastocoel cavity expansion relative to the zona pellucida, the compactness of the inner cell mass, and the density of the trophectoderm. Accurate measurement of these parameters requires segmenting human blastocyst images to delineate the three regions. Manual identification of these regions is challenging, time-consuming, and subjective, often leading to variability between experts. Recently, many efforts have been made to automate embryo segmentation, utilizing both classical and deep learning-based approaches (Filho et al., 2012; Saeedi et al., 2017; Rad et al., 2019; Muhammad et al., 2022). Despite its importance, embryo segmentation still presents several challenges due to significant variability in the scale, shape, position, and orientation of blastocyst components. Convolutional Neural Networks (CNNs) have emerged as state-of-the-art tools in medical image segmentation, offering remarkable performance in various applications. However, conventional CNN architectures face certain limitations when applied to this task. One key issue is their limited spatial awareness, particularly for flexible structures with varying shapes and positions, as CNNs rely on shared weights in convolutional layers. This limitation can affect accurate segmentation, especially when dealing with complex, non-rigid structures such as the inner cell mass, trophectoderm, and zona pellucida. Moreover, the high dimensionality of feature maps in CNNs can lead to redundancy, reducing computational efficiency and increasing processing time. Another significant challenge is the lack of interpretability in CNN decision-making. While CNNs excel at extracting features, their black-box nature makes it difficult to understand why certain decisions are made. This lack of explainability is a major drawback in clinical settings, where transparent decision-making is crucial for medical professionals to trust and act on the results.

To address these challenges, this work focuses on improving human embryo segmentation. As a critical step in the IVF process, embryo segmentation enables clinicians to assess embryo viability by delineating key regions, including Inner Cell Mass, Zona Pellucida, and Trophectoderm. Accurate segmentation of these regions is essential for identifying embryos with the highest implantation potential, thereby improving IVF success rates. Building on recent advancements in deep learning, this study proposes a novel approach to tackle the dual challenges of segmentation accuracy and model interpretability.
The main contributions of this work are as follows:

- We integrate channel attention mechanisms with a ResNet50 backbone in an encoder-decoder convolutional neural network. This architecture enhances the model's ability to detect key human blastocyst components while providing interpretable insights into the decision-making process.
- The proposed approach is evaluated using a publicly available blastocyst segmentation dataset. Its performance is compared to state-of-the-art methods and a baseline U-Net model augmented with the post-hoc explainability technique Grad-CAM.
- This work addresses the dual challenges of accuracy and interpretability in deep learning models for embryo segmentation. By providing a framework that ensures reliable segmentation while enhancing model transparency, it contributes to better healthcare outcomes and fosters greater trust in AI-driven medical solutions.

2 EXPLAINABLE AI FOR IMAGE SEGMENTATION

Segmentation involves the automatic identification and delineation of specific objects or regions within an image. The integration of explainable artificial intelligence (XAI) into image segmentation is gaining attention for its potential to enhance the fairness and reliability of AI models, particularly in sensitive domains like healthcare (Saranya and Subhashini, 2023; Gunashekar et al., 2022). In clinical applications, interpretable segmentation models are essential to build trust, improve communication with patients, and support iterative refinements to achieve more precise and accurate results. This section explores XAI methods designed to clarify the decision-making processes of segmentation models, a critical step toward enhancing their applicability in clinical contexts and advancing research efforts.

2.1 Ad-Hoc Explainability Methods

Ad-hoc explainability methods are external techniques applied to pre-existing AI models to provide insight into their decision-making processes. These methods are particularly valuable for complex models, such as deep learning architectures, which inherently lack transparency. Commonly used ad hoc techniques for image segmentation include SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM). SHAP assigns contribution scores to input features, quantifying their influence on the model's predictions. By leveraging game theory, SHAP ensures fair attribution of prediction credit, providing a detailed understanding of the importance of the feature. LIME generates interpretable surrogate models to approximate the behavior of a complex model around a specific prediction. This helps explain individual predictions, especially in ambiguous or error-prone cases. The Grad-CAM approach (Selvaraju et al., 2017; Vinogradova et al., 2020) utilizes gradients from the final convolutional layer to highlight image regions that most influence the model's predictions, providing intuitive visual explanations. Its enhanced variant, Grad-CAM++, incorporates information from all convolutional layers, refining the localization of discriminative regions and delivering more precise explanations.

2.2 Self-Explainable Methods

Unlike ad-hoc methods, self-explainable models are designed with interpretability as an integral part of their architecture. These models inherently integrate explainability into their structure, making their decision-making processes clear. One prominent example is the ProtoSeg method, proposed by Sacha *et al.* (Sacha et al., 2023), which uses prototype-based learning to assign pixels to classes by comparing image regions to prototypes, which are characteristic examples of each class. This approach naturally enhances interpretability, as segmentation decisions are explicitly linked to these prototypes. Attention-based models leverage attention mechanisms to focus on relevant regions within an image during segmentation (Gu et al., 2020; Zhao et al., 2021). By assigning importance scores to features, they generate interpretable attention maps, simultaneously improving segmentation accuracy and transparency.

The choice between ad hoc and self-explainable methods depends on the specific requirements of the applications. Ad-hoc methods are suitable for providing detailed and flexible analysis, making them ideal for in-depth evaluations of existing models. However, they often require technical expertise and computational resources. In contrast, self-explainable methods prioritize intuitive interpretation and computational efficiency, making them well-suited for resource-constrained environments or users with limited technical expertise.

3 METHODOLOGY

3.1 Model Architecture

3.1.1 Overview

The proposed Residual Attention Network (ResA-Net), inspired by the method introduced in (Gu et al., 2020), integrates comprehensive attention mechanisms for efficient and interpretable image segmentation. While retaining the general structure of the original model, we enhanced its backbone by replacing it with ResNet50 to leverage its superior feature extraction capabilities. Specialized attention modules are incorporated to guide segmentation across spatial, channel, and scale dimensions simultaneously. The architecture comprises four spatial attention modules (SA1-SA4), four channel attention modules (CA1-CA4), and a scale attention module (LA), each contributing to precise and context-aware segmentation.

ResA-Net adopts an encoder-decoder architecture. The encoder processes input images through convolutional layers and max-pooling operations, reducing spatial dimensions while capturing high-level and multi-scale features. These features are passed to a central layer, bridging the encoder and decoder, before being fed into the decoder to reconstruct the segmentation output. The decoder employs bilinear interpolation and concatenation to upsample feature maps to the original resolution while effectively combining information from multiple scales. Finally, a 1×1 convolutional layer reduces feature map dimensions, and a class-specific convolutional layer generates probability maps for each segmentation class. A softmax layer converts these probabilities into the final segmentation output by assigning each pixel to the class with the highest likelihood. An overview of the proposed method is presented in Figure 1.

While achieving accurate segmentation is essential, our focus extends beyond the segmentation task to understanding the model's decision-making process. To this end, we emphasize the visualization of attention maps generated by ResA-Net, which highlight influential regions in the input image. These

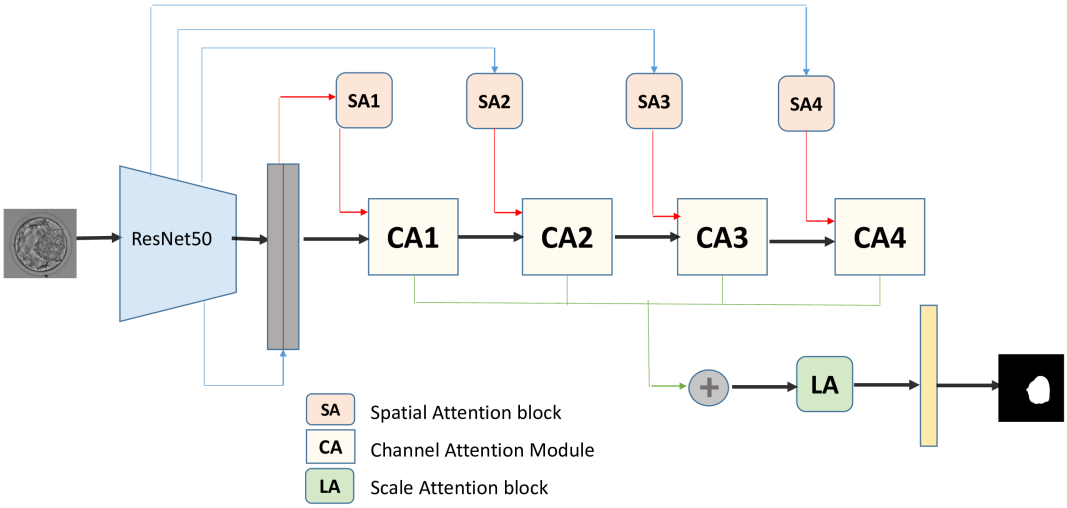

Figure 1: Overview of the proposed ResA-Net model.

visualizations offer valuable insights into how the model makes its decisions, enhancing transparency and interpretability by clearly showcasing the regions and features that guide its segmentation outcomes.

3.1.2 Attention Mechanisms in ResA-Net

ResA-Net employs three attention mechanisms Spatial Attention Modules (SA), Channel Attention Modules (CA), and a Scale Attention Module (LA).

Spatial Attention: Inspired by the non-local network and Attention Gates (AG), ResA-Net uses four spatial attention blocks (SA1-SA4) to learn attention maps across different resolution levels. These maps emphasize relevant regions of the input image while suppressing noise, enhancing segmentation accuracy. At the lowest resolution level (SA1), a non-local block captures global context, while at higher levels (SA2-SA4), attention gates focus on key areas within the image. Low-level spatial attention features from the encoder are concatenated with high-level decoder features to enhance segmentation through complementary information.

Channel Attention: Channel attention modules automatically identify and amplify relevant feature channels while suppressing irrelevant ones. This mechanism ensures that the model focuses on semantically meaningful features crucial for accurate segmentation.

Scale Attention: To handle variations in object scale, the scale attention module assigns adaptive weights to features at different scales. This allows the network to prioritize features most relevant to the specific image, improving segmentation performance.

By combining these mechanisms, ResA-Net not only achieves high segmentation accuracy but also provides interpretable outputs, offering insights into the decision-making process and enabling its application in sensitive fields such as healthcare.

3.2 Dataset

The dataset employed in this paper for segmenting human embryo images is the unique publicly available dataset introduced by Saaedi *et al.* (Saeedi et al., 2017). It comprises 235 blastocyst images from patients treated at the Pacific Center of Reproductive Medicine in Canada between 2012 and 2016. Each image was manually annotated by expert embryologists and provided with ground truth binary masks relating to ICM, TE, and ZP regions. Using this dataset, we employed the proposed model architecture to train three distinct configurations, each focused on segmenting a specific region of interest. This work adopts a single-class segmentation approach to individually identify the ZP, ICM, and TE regions within blastocyst images. By generating explicit segmentation labels, the model facilitates both accurate segmentation and enhanced interpretability analysis.

4 EXPERIMENTAL RESULTS

4.1 Implementation Details

A consistent preprocessing pipeline was applied to all three one-class segmentation configurations to ensure uniformity and efficiency. All images and their corre-

sponding ground truth masks were resized to a fixed resolution of 224×300, to ensure consistent input dimensions for the model. Both the images and masks were then converted to `.npy` format to facilitate efficient storage and handling during training.

To manage the dataset, we implemented a custom dataset class using PyTorch's `Dataset` module. This class was specifically designed to pair image samples with their corresponding labels, with each dataset entry represented as tensors loaded from the preprocessed `.npy` files. The dataset was divided into training (85%), and testing (15%) subsets, ensuring a balanced evaluation of the model's performance. Data loading was performed using PyTorch's `DataLoader`, enabling efficient batching and streamlined access to the dataset. A batch size of 16 images was selected to balance memory constraints and training efficiency. The Adam optimizer was employed with a learning rate of 10^{-4} and a weight decay of 10^{-8}, to optimize the model parameters. The Dice loss function was utilized to evaluate segmentation performance, as it effectively handles the class imbalance inherent in medical imaging datasets. Each configuration was trained for 100 epochs to ensure convergence and optimal performance. Training was conducted locally on a desktop PC equipped with an NVIDIA GeForce GTX 1650 graphics card, providing sufficient computational power for the segmentation task.

4.2 Evaluation

For the performance evaluation of the proposed model, we used the Intersection over Union (IoU), also known as the Jaccard Index, and the F1 score. The Jaccard Index is the ratio of the intersection to the union of the segmentation result and the ground truth. The F1 score is defined as twice the area of overlap between the segmentation result and the ground truth, divided by the sum of the areas of both. Table 1 presents a quantitative comparison of the segmentation performance between the proposed method, the baseline U-Net model and recent state-of-the art methods. U-Net (Ronneberger et al., 2015) has demonstrated its effectivenes in segmenting anatomical structures across various imaging modalities, including MRI, CT scans, and microscopy. BlasNet (Rad et al., 2019) enhances segmentation by incorporating multi-scale global contextual information through a cascaded atrous pyramid pooling module and reproducing feature resolution using dense progressive sub-pixel upsampling. MASS-Net (Muhammad et al., 2022) leverages depth-wise concatenation to combine spatial information at multiple scales, enabling the simultaneous detection of blastocyst components. ECS-Net (Mushtaq et al., 2022) employs dual streams: base convolutional and depth-wise separable convolutional blocks, densely concatenated to enrich features for improved segmentation. FSBS-Net (Ishaq et al., 2023) introduces feature supplementation across scales and employs ascending channel convolutional blocks (ACCB) to refine blastocyst segmentation with minimal computational overhead, leveraging skip connections to integrate features from shallow and deep layers.

Based on the Intersection over Union (IoU) metric, the ResA-Net model consistently outperforms state-of-the-art models in segmenting the ZP and TE regions. While its performance on the ICM region is slightly below that of some state-of-the-art models, ResA-Net achieves outstanding results in terms of the mean IoU across all three regions. This demonstrates the model's overall effectiveness and robustness in segmentation tasks.

4.3 Models Interpretability

4.3.1 Visualization with Grad-CAM for U-Net

To better understand the areas of interest that guide the U-Net model's segmentation decisions, we used Grad-CAM to generate heatmaps for each segmentation class. These heatmaps were then combined into a single heatmap that highlights the regions the model focuses on during the segmentation process. Grad-CAM was applied to the last convolutional layer of the U-Net model, providing insight into how the model attends to different regions of the image. For the ZP region, the heatmap shows clear focus on specific areas within the embryo's perimeter, which is the most straightforward to segment. In contrast, for the ICM region, the model appears to focus on a single, central entity within the embryo. This suggests that the model identifies a distinct, central structure to guide segmentation. The Trophectoderm (TE) region, which has the lowest segmentation metric scores, is the most challenging to identify. The Grad-CAM heatmap reveals that the model attempts to identify multiple entities within the inner perimeter of the embryo, which later converge to form the TE region. Figure 2 presents examples of the combined Grad-CAM outputs applied to the final layers of the three U-Net configurations, illustrating the model's focus for each segmentation class.

4.3.2 Visualization of ResA-Net Attention Maps

Figure 3 presents the attention maps generated by the ResA-Net model for each segmentation configuration (ICM, ZP, TE), derived from the last image in the fi-

Table 1: Comparative analysis of proposed method with state-of-the-art models in terms of IoU and F1 score.

Model	ZP		TE		ICM		Mean
	IoU	F1	IoU	F1	IoU	F1	IoU
U-Net Baseline	79.10	88.33	71.66	83.49	76.33	86.57	75.69
BlastNet (Rad et al., 2019)	81.15	-	76.52	-	81.07	-	79.58
MASS-Net (Muhammad et al., 2022)	84.69	-	79.08	-	**85.88**	-	83.22
ECS-Net (Mushtaq et al., 2022)	85.34	-	78.43	-	85.26	-	83.01
FSBS-Net (Ishaq et al., 2023)	85.80	92.29	80.17	**88.90**	85.55	**92.0**	**83.84**
ResA-Net	**86.87**	**92.72**	**81.10**	88.88	83.09	90.09	83.68

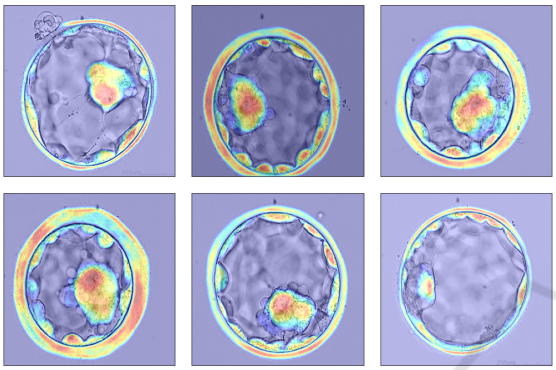

Figure 2: GradCAM heatmaps for U-Net model.

nal batch of the training dataset. These maps offer valuable insights into the model's segmentation process for each class. In the case of the Zona Pellucida one-class segmentation configuration, the attention maps reveal a precise and systematic segmentation process. The first map clearly delineates the ZP from its surrounding background, while subsequent maps progressively refine the segmentation by emphasizing key feature channels, particularly highlighting the boundary between the ZP and TE regions. This progression helps accurately define the contours separating these regions, showcasing the model's ability to focus on the critical features needed for precise segmentation.

For the Inner Cell Mass one-class configuration, the attention maps demonstrate how the model identifies the relevant regions within the blastocyst. The initial map shows the model's broad understanding of the entire blastocyst, detecting it as a whole. The second map narrows in on the ICM, accurately localizing this important region. This sequential refinement indicates the model's capacity to differentiate between the various components of the blastocyst, enhancing segmentation precision. However, a slight overlap is observed between the ICM and TE regions, where some TE pixels are also highlighted. While this minor ambiguity does not significantly affect the overall performance, it underscores the challenges of distinguishing adjacent regions.

In contrast, the attention maps for the Trophectoderm one-class configuration reveal challenges in segmentation. Unlike the ZP and ICM models, the TE model struggles to isolate the TE region. The initial attention maps fail to clearly differentiate the TE from the background or the blastocyst structure. Despite relatively high segmentation metrics for the TE region, these scores are lower compared to those for ZP and ICM, suggesting that this particular input image may not fully represent the model's performance across the entire dataset.

5 CONCLUSION

In IVF procedures, accurate segmentation is essential for assessing embryo viability. By dividing the embryo into distinct regions, such as the Inner Cell Mass, Zona Pellucida, and Trophectoderm, clinicians can better identify embryos suitable for implantation, ultimately improving the chances of a successful pregnancy. This paper introduces the development of three one-class segmentation models for embryo image analysis, incorporating attention mechanisms with a ResNet50 backbone in an encoder-decoder architecture. Experimental results demonstrate that the integration of attention mechanisms enhances the model's ability to focus on relevant regions, leading to improved segmentation performance.

To further enhance segmentation results, future work could focus on refining the current models by incorporating additional contextual information, which would improve the differentiation of closely related regions. Additionally, exploring more advanced attention mechanisms and hybrid architectures, combining CNNs with transformer models, could capture fine-grained details more effectively, ultimately boosting performance.

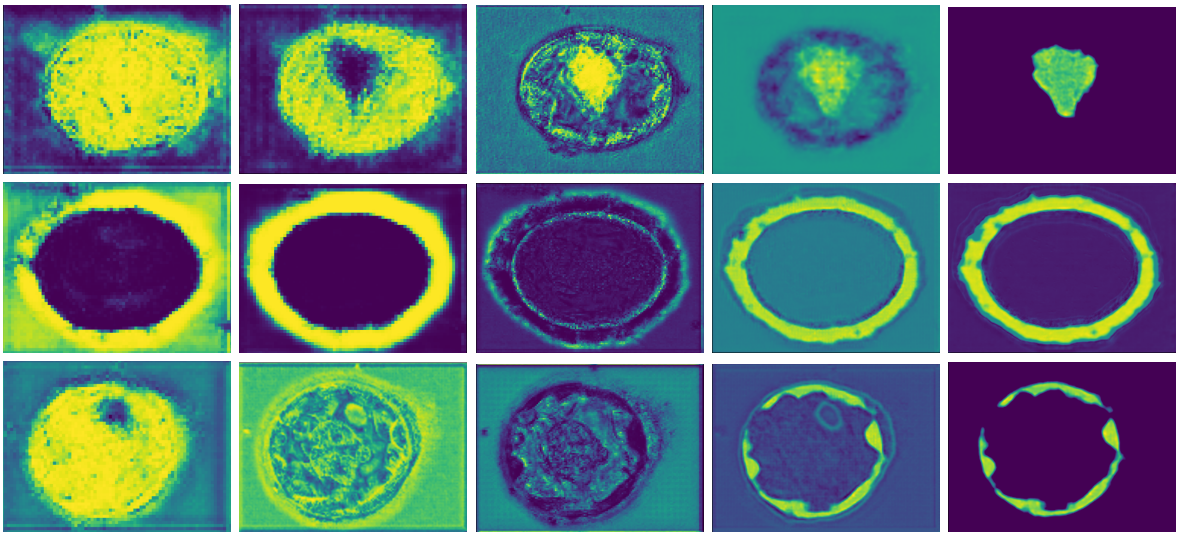

Figure 3: Attention Heatmaps Generated by the ResA-Net Model.

REFERENCES

Dennis, S. J., Thomas, M. A., Williams, D. B., and Robins, J. C. (2006). Embryo morphology score on day 3 is predictive of implantation and live birth rates. *Journal of assisted reproduction and genetics*, 23(4).

Filho, E., Noble, J., Poli, M., Griffiths, T., Emerson, G., and Wells, D. (2012). A method for semi-automatic grading of human blastocyst microscope images. *Human reproduction (Oxford, England)*, 27:2641–2648.

Gardner, D. K. and Schoolcraft, W. B. (1999). Culture and transfer of human blastocysts. *Current Opinion in Obstetrics and Gynaecology*, 11(3):307–311.

Goodman, L. R., Goldberg, J., Falcone, T., Austin, C., and Desai, N. (2016). Does the addition of time-lapse morphokinetics in the selection of embryos for transfer improve pregnancy rates? a randomized controlled trial. *Journal of Fertility and Sterility*, 105(2).

Gu, R., Wang, G., Song, T., Huang, R., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., and Zhang, S. (2020). Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE transactions on medical imaging*, 40(2):699–711.

Gunashekar, D. D., Bielak, L., Hägele, L., Oerther, B., Benndorf, M., Grosu, A., Brox, T., Zamboglou, C., and Bock, M. (2022). Explainable ai for cnn-based prostate tumor segmentation in multi parametric mri correlated to whole mount histopathology. *Radiation Oncology*, 17(1):65.

Ishaq, M., Raza, S., Rehar, H., Abadeen, S., Hussain, D., Naqvi, R., and Lee, S. W. (2023). Assisting the human embryo viability assessment by deep learning for in vitro fertilization. *Mathematics*, 11(9):2023.

Kovacs, P. (2016). Time-lapse embryoscopy: Do we have an efficacious algorithm for embryo selection? *Journal of Reproductive Biotechnology and Fertility*, 5.

Kuhnt, A. K. and Passet-Wittig, J. (2022). Families formed through assisted reproductive technology: Causes, experiences, and consequences in an international context. In *Reprod Biomed Soc Online*, number 14, pages 289–296.

Muhammad, A., Adnan, H., Woon, C. S., Hwan, K. Y., and Ryoung, P. K. (2022). Human blastocyst components detection using multiscale aggregation semantic segmentation network for embryonic analysis. *Biomedicines*, 10(7):1717.

Mushtaq, A., Mumtaz, M., Raza, A., Salem, N., and Yasir, M. N. (2022). Artificial intelligence-based detection of human embryo components for assisted reproduction by in vitro fertilization. *Sensors*, 22(19):7418.

Rad, R. M., Saeedi, P., Au, J., and Havelock, J. (2019). Blast-net: Semantic segmentation of human blastocyst components via cascaded atrous pyramid and dense progressive upsampling. In *IEEE International Conference on Image Processing (ICIP)*, pages 1865–1869.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

Sacha, M., Rymarczyk, D., Struski, L., Tabor, J., and Zieliński, B. (2023). Protoseg: Interpretable semantic segmentation with prototypical parts. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1481–1492.

Saeedi, P., Yee, D., Au, J., and Havelock, J. (2017). Automatic identification of human blastocyst components via texture. *IEEE Transactions on Biomedical Engineering*, 64(12):2968–2978.

Saranya, A. and Subhashini, R. (2023). A systematic re-

view of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision analytics journal*, page 100230.

Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.

Shulman, A., Ben-Nun, I., Y. Ghetler, H. K., Shilon, M., and Beyth, Y. (1993). Relationship between embryo morphology and implantation rate after in vitro fertilization treatment in conception cycles. *Fertility and sterility*, 60(1).

Vinogradova, K., Dibrov, A., and Myers, G. (2020). Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(10):13943–13944.

Zhao, Q., Liu, J., Li, Y., and Zhang, H. (2021). Semantic segmentation with attention mechanism for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, PP:1–13.